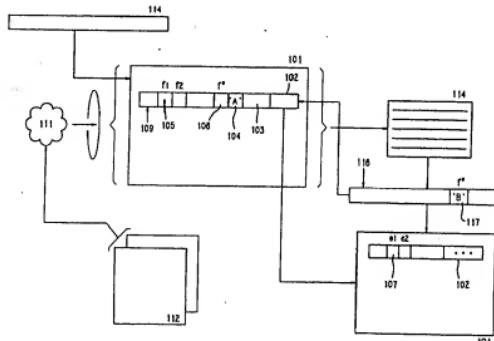




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : G06F 15/18, 15/40		A2	(11) International Publication Number: WO 93/21587
			(43) International Publication Date: 28 October 1993 (28.10.93)
<p>(21) International Application Number: PCT/US93/03558</p> <p>(22) International Filing Date: 14 April 1993 (14.04.93)</p> <p>(30) Priority data: 07/869,935 15 April 1992 (15.04.92) US</p> <p>(71) Applicant: INFERENCE CORPORATION [US/US]; 550 N. Continental Blvd., El Segundo, CA 90245 (US).</p> <p>(72) Inventor: ALLEN, Bradley, Paul ; 850 Loma Drive, #B, Hermosa Beach, CA 90254 (US).</p> <p>(74) Agents: SWERNOFSKY, Steven, A. et al.; Lyon & Lyon, 611 West Sixth Street, 34th Floor, Los Angeles, CA 90017 (US).</p>		<p>(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>	
<p>(54) Title: MACHINE LEARNING WITH A RELATIONAL DATABASE</p> <p>(57) Abstract</p> <p>A machine learning system which operates in conjunction with a relational database. The system may (1) examine a selected entry in the database, (2) query the database for a set of entries which are representative of the selected entry, and (3) predict a value for one or more fields of the selected entry in response to the set of representative entries. The system may perform these steps repeatedly, and may evaluate each entry and record an indication of accuracy or utility (or other values) of that entry for predicting one or more fields. The system may also implement a case-based reasoning system, or an autonomous learning system, with a relational database. A system for error-checking and correlating entries and fields in a relational database. The predicted values for one or more fields of the selected entry may be compared with the actual values. The system may note field values which differ too much from predicted as possibly erroneous (or at least as data which should be checked). The system may "fill in" fields with the predicted values if actual values are missing or distrusted. Occasional or periodic error-checking and selective replacement of erroneous data may provide a self-repairing database. The system may also note fields which are easy to predict as redundant, may note tuples of fields which are strongly correlated as causally related, or may note fields are difficult to predict as requiring other data for good prediction.</p>			



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Mongo	TC	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

DESCRIPTIONMachine Learning With A Relational Database
Cross-Reference To Related Application

This application is a continuation-in-part of copending application Serial No. 07/ 664,561, filed March 4, 1991 in the name of inventors Bradley P. Allen and S. Daniel Lee and titled "CASE-BASED REASONING SYSTEM", 5 hereby incorporated by reference as if fully set forth herein.

BACKGROUND OF THE INVENTION1. Field of the Invention

This invention relates to machine learning. More 10 specifically, this invention relates to a machine learning system which uses a relational database.

2. Description of Related Art

While computers are capable of tremendous processing power, their ability to use that processing power for 15 reasoning about complex problems has so far been limited. Generally, before a computer can be used to address a complex problem, such as one which requires the attention of a human expert, it has been necessary to distill the knowledge of that expert into a set of inferential rules 20 (a "rule base") which allow an automated processor to reason in a limited field of application. While this method has been effective in some cases, it has the natural drawback that it often requires a substantial amount of time and effort, by both computer software 25 engineers and experts in the particular field of application, to produce a useful product.

Moreover, rule-based systems of this type present a difficult programming task. Unlike more prosaic programming tasks, constructing a rule base is sometimes 30 counterintuitive, and may be beyond the ability of many

application programmers. And once a rule-based system has been constructed based on the knowledge of a human expert, it may be difficult to accommodate changes in the field of operation in which the processor must operate. Such 5 changes might comprise advances in knowledge about the application field, additional tasks which are intended for the processor, or changes in or discoveries about the scope of the application field.

One proposed method of the prior art is to build 10 automated reasoning systems which operate by reference to a set of exemplar cases (a "case base"), to which the facts of a particular situation (the "problem") may be matched. The processor may then perform the same action for the problem as in the exemplar case. While this 15 proposal has been well-received, case-based systems of this type may still require a substantial amount of human effort to identify exemplar cases and present a processor with sufficient information that cases may be matched and acted upon. For example, it may be necessary to deduce or 20 supply extensive information about a complex environment so as to determine a preferred set of exemplar cases.

A parent copending application, Serial No. 07/664,561, filed March 4, 1991, discloses inventions in which a case-based reasoning system is smoothly integrated 25 into a rule-based reasoning system, and in which an automated reasoning system may dynamically adapt a case base to problems which it encounters. An aspect of the invention disclosed in that application also includes a technique in which a system may be set to work with a 30 limited case base, and may solicit human advice for treatment of new problems which are not already well-treated by the case base, thus learning how to do its job on a dynamic basis.

Another copending application, Serial No. 35 _____, Lyon & Lyon Docket No. 193/304, filed the same day as this application, discloses inventions in which an automated reasoning system may dynamically create

its own case base in response to problems which it encounters, thus learning how to do its job on a dynamic basis and without substantial human intervention, or at least with only occasional human intervention. In the 5 inventions disclosed therein, an automated reasoning system may also operate autonomously in a complex environment, possibly with external intervention such as positive or negative reinforcing stimuli. The external stimuli might be in response to a result of the system's 10 attempts to manipulate its environment, or might be provided by an external agent, such as a human operator.

In some of the many fields which relate to computing, one interesting development has been the increasing processing power which has been applied to databases. 15 Many computer systems which have nothing to do with learning or reasoning systems have substantial databases, and may execute software which manipulates or queries those databases repeatedly. The software may require intensive searching of a large memory under complex search 20 conditions. Accordingly, computer systems have been developed which can search databases at high speed, and in particular, can search relational databases at high speed using SQL, a standard query language for relational databases.

25 It would be advantageous if a machine learning system could operate in conjunction with a relational database system, and particularly advantageous if a machine learning system could operate in conjunction with a relational database system with an SQL interface. This 30 would allow the machine learning system to use the high-speed searching power of these computer systems, and would allow the machine learning system to be smoothly integrated into computer systems which have relational databases, even if those databases were not designed to 35 work with learning or reasoning systems of any kind. Accordingly, it is an object of the invention to provide

a machine learning system which may operate with a relational database.

Summary of the Invention

The invention provides a machine learning system which operates in conjunction with a relational database. The machine learning system may (1) examine a selected entry in the database, (2) query the database for a set of entries which are representative of the selected entry, and (3) predict a value for one or more fields of the selected entry in response to the set of representative entries. In a preferred embodiment, the system may perform these steps repeatedly, and may evaluate each entry and record an indication of accuracy or utility (or other values) of that entry for predicting one or more fields.

The invention also provides an implementation of a case-based-like reasoning system with a relational database. In such a reasoning system, the entries of the database may correspond generally to cases in a case-based reasoning system, the fields may correspond generally to features in a case-based reasoning system, searching the database may correspond generally to matching cases in a case base, and predicting one or more fields may correspond generally to selecting a case to use in a case-based reasoning system. Evaluating each entry may correspond generally to evaluating accuracy or utility (or other values) of cases for prescribing a correct action to take.

The invention also provides an implementation of an autonomous learning system with a relational database. In a preferred embodiment, the machine learning system may implement an autonomous learning software agent, like that disclosed in copending application Serial No. _____, Lyon & Lyon Docket No. 193/304. In such a software agent, new entries in the database may be generated, deleted or modified by means of techniques

which correspond generally to those by which new cases are generated, deleted or modified, as shown in that copending application, or as shown in parent copending application Serial No. _____, Lyon & Lyon Docket No. 5 193/108.

The invention also provides a system for error-checking and correlating entries and fields in a relational database. The predicted values for one or more fields of the selected entry may be compared with the 10 actual values. The system may note field values which differ too much from predicted as possibly erroneous (or at least as data which should be checked). Alternatively, the system may "fill in" fields with the predicted values if actual values are missing or distrusted. Occasional or 15 periodic error-checking and selective replacement of erroneous data may provide a self-repairing database. Moreover, the system may also note fields whose values which are easy to predict as possibly redundant, may note tuples of fields which are strongly correlated as possibly 20 causally related, or may note fields whose values are difficult to predict as possibly requiring other data for good prediction.

Brief Description of the Drawings

Figure 1A shows a data flow diagram of a method of 25 machine learning with a relational database. Figure 1B shows a process flow diagram of a method of machine learning with a relational database.

Figure 2A shows a data flow diagram of a method of 30 cluster recognition with a relational database. Figure 2B shows a process flow diagram of a method of cluster recognition with a relational database.

Appendix A shows an example software environment and autonomous agent for distinguishing between classes of irises.

Description of the Preferred Embodiment

An embodiment of this invention may be used together with inventions which are disclosed in a copending application titled "AUTONOMOUS LEARNING AND REASONING AGENT", application Serial No. _____, Lyon & Lyon Docket No. 193/304, filed the same day in the name of the same inventor, hereby incorporated by reference as if fully set forth herein.

Figure 1A shows a data flow diagram of a method of machine learning with a relational database. Figure 1B shows a process flow diagram of a method of machine learning with a relational database.

A relational database 101 may comprise a set of records 102 and a set of fields 103, as is well known in the art. Each field 103 in each record 102 may comprise a value 104, such as a numeric value, a string value, or a value with another data type, as is well known in the art. Relational databases are more fully described in "Principles of Database Systems", by Jeffery D. Ullman, published by Computer Science Press, hereby incorporated by reference as if fully set forth herein.

The database 101 may comprise at least one feature field 105 f1, f2, . . . fn, at least one predicted field 106 f*, and a set of evaluation fields 107 e1, e2, . . . en. In a preferred embodiment, the feature fields 105, predicted fields 106, and evaluation fields 107 are all found in a single database 101. However, it would be clear to one of ordinary skill in the art, after perusal of the specification, drawings and claims herein, that the database 101 may be organized in a variety of different ways consistent with the art of relational databases. For example, the evaluation fields 107 may form separate records 102 in a second database 101, correlated with the first database 101 by a set of record identifiers or by some similar technique, as is well known in the art. It would also be clear that many different ways of

organization would be workable, and are within the scope and spirit of the invention.

In a record-designation step 108, a single record 102 in the database 101 may be designated as a selected record 5 109.

In a query-composition step 110, the selected record 109 may be examined and a database query or search designation 111 may be composed for records 102 which are "similar". In a preferred embodiment, the search 10 designation 111 may be specified in the SQL language, as is well known in the art. However, it would be clear to one of ordinary skill in the art, after perusal of the specification, drawings and claims herein, that other query languages or techniques for designating searches may 15 also be used, that such other languages and techniques would be workable, and are within the scope and spirit of the invention. Techniques for manipulating and querying databases using the SQL language are more fully described in "SQL Language Reference Manual (Version 5.1)", 20 published by Oracle Corporation, hereby incorporated by reference as if fully set forth herein.

A set of similarity tables 112 may be maintained which indicate what records 102 are regarded as similar to the selected record 109. For example, the similarity 25 tables 112 may indicate that a record 102 is similar to the selected record 109 if its value for the feature field 105 f1 is within 0.1 numeric units, its value for the feature field 105 f2 is within 0.2 numeric units, its value for the feature field 105 f3 shares at least 3 30 common characters of text, and so on. Also, techniques for evaluating similarity such as like those disclosed in parent copending application Serial No. _____, Lyon & Lyon Docket No. 193/108, may be used. It would be clear to one of ordinary skill in the art, after perusal 35 of the specification, drawings and claims herein, that various different techniques for measuring similarity may

be used, that such different techniques would be workable, and are within the scope and spirit of the invention.

In a query-application step 113, the search designation 111 may be applied to the database 101, to 5 produce a search set 114, comprising a set of records 102 which meet the search designation 111.

In a predictor-selection step 115, one or more evaluation fields 107 of the records 102 in the search set 114 may be examined, and a predictive record 116 may be 10 chosen for one or more predicted fields 106. Techniques such as like those used in the selector module of copending application Serial No. _____, Lyon & Lyon Docket No. 193/304, may be used to chose the predictive record 116. However, it would be clear to one 15 of ordinary skill in the art, after perusal of the specification, drawings and claims herein, that other and further techniques might also be used, that such other and further techniques would be workable, and are within the scope and spirit of the invention. A predicted value 117 20 for the predicted field 106 f^* is the value for f^* which is found in the predictive record 116.

In a evaluation-update step 118, the predicted value 117 from the predictive record 116 may be compared with the actual value 104 found in the selected record 109, and 25 the evaluation fields 107 of the predictive record 116 may be updated accordingly. In a preferred embodiment, the evaluation fields 107 may comprise fields for "times used", "times correct", "accuracy", "utility", and other evaluative measures such as like those disclosed in 30 copending application Serial No. _____, Lyon & Lyon Docket No. 193/304.

Each record 102 of the database 101 may be designated as the selected record 109, and the process of designating a selected record 109, composing and applying a search 35 designation 111, choosing a predictive record 116, comparing with the selected record 109 and updating the predictive record 116, may be performed repeatedly. This

causes values found in the evaluation fields 107 of the database 101 to reach an equilibrium state, in which they accurately represent, or at least closely approximate, the true accuracy and utility of the predictive value of each 5 record 102. As used herein, a "predictive" database is a database 101 which is substantially in such a state, and "predictive relaxation" is the technique of repeatedly updating the evaluation fields 107 which helps make a database 101 predictive.

10 When a record 102 is added to, deleted from, or modified in the predictive database 101, predictive relaxation may be repeated so as to maintain the database 101 predictive. In a preferred embodiment, predictive relaxation may be performed logically in parallel with 15 other database operations such as adding, deleting or modifying records 102, so that the database 101 is maintained predictive even as it changes. The database 101 may also be maintained predictive while the similarity tables 112 are altered.

20 When a new record 119 is added to the predictive database 101, a set of predicted values 117 for one or more of its fields 103 may be determined, and the predicted values 117 compared with the actual values 104 from the new record 119. The feature fields 105 and the 25 predicted fields 106 may overlap, i.e., one or more feature fields 105 may also be predicted fields 106, so that any field 103 may be predicted. When more than one field is predicted, the evaluation fields 107 for predicting that field 103 may differ from the evaluation 30 fields 107 for predicting another field 103.

If one or more fields 103 in the new record 119 have no defined values 104, the values 104 for those fields 103 may be filled in by predicting them. Thus, if field 103 f* is missing its value 104, the f* value 104 from the 35 predictive record 116 may be inserted. Such values 104 might be inserted when the new record 119 is added, or at a later time.

If one or more fields 103 in the new record 119 have values 104 which differ substantially from predicted, an alarm signal may be generated to indicate that such values 104 are erroneous, or at least should be checked. Such an 5 alarm signal might be generated when the new record 119 is added, or might be generated occasionally as the database 101 is maintained predictive.

If an alarm signal indicates that values 104 in the new record 119 are possibly erroneous, or the values 104 10 are otherwise distrusted, the distrusted values 104 might be replaced with the predictive values 104. Replacing such values 104 might occur when the new record 119 is added, or might occur occasionally as the database 101 is maintained predictive. Alternatively, if the lack of an 15 alarm signal indicates that values 104 in the new record 119 are probably correct, and the values 104 are otherwise distrusted, the distrusted values 104 might be marked trustworthy. The database 101 may be maintained predictive in logical parallel with selectively replacing 20 erroneous values 104, thus providing a database 101 which self-repairs any erroneous values 104 which are introduced in the course of adding, deleting, or modifying records 102.

Figure 2A shows a data flow diagram of a method of 25 cluster recognition with a relational database. Figure 2B shows a process flow diagram of a method of cluster recognition with a relational database.

The machine learning system may include a technique for cluster recognition. The machine learning system may 30 determine new records 102 for each feature field 105 which represent clusters 201 of values 104 for that feature field 105. The new records 102 may be added to the database 101, or may be used to create a second database 101 which incorporates essentially the same information. 35 Knowledge about clusters 201 may also be used in the similarity tables 112, for example, to indicate that a value 104 for a feature field 105 of a record 102 is

similar to a value 104 for the same feature field 105 of the selected record 109 if both values 104 are in the same cluster 201 for that feature field 105.

5 In a feature-selection step 202, a feature field 105 is selected for cluster recognition.

10 In a cluster-setup step 203, an initial cluster-count of clusters 201 is set. The initial cluster-count may be selected arbitrarily or randomly, by known statistical methods, or might be provided by an external agent, such as a human operator. For example, in a preferred embodiment, the initial cluster-count may always be set to four clusters 201.

15 In an alternative embodiment, methods shown in copending application Serial No. _____, Lyon & Lyon Docket No. 193/304, which are applicable to cases, may also be applied for setting the cluster-count. For example, the cluster-count may be set to maximize "accuracy", "utility", and other evaluative measures such as like those disclosed in that application, of the 20 resulting set of clusters 201.

25 In a cluster-centroid step 204, the range 205 of possible values 104 for the feature field 105 may be divided into subranges 206, one per cluster 201, and for each cluster 201, a cluster centroid 207 is selected. In a preferred embodiment, each cluster centroid 207 may be selected arbitrarily or randomly within its subrange 206, but the cluster centroid 207 might be selected by a known averaging technique (such as the averaging technique used in the cluster-averaging step 209 herein, or provided by 30 an external agent, such as a human operator.

35 In a cluster-query step 208, a single cluster 201 may be selected and the database 101 may be interrogated for members of that cluster 201. This step includes composing and applying a search designation 111 for the database 101, in similar manner as shown with respect to figures 1A and 1B.

In a cluster-averaging step 209, an averaging technique is applied to the members of the cluster 201, in response to which a target centroid 210 may be determined. The old cluster centroid 207 may be adjusted toward the 5 target centroid 210, e.g., by an exponential decay technique. In a preferred embodiment, a predetermined fraction, such as 80%, of the difference between the old cluster centroid 207 and the target centroid 210, is added to the old cluster centroid 207 to determine a new cluster 10 centroid 207.

The cluster-query step 208 and the cluster-averaging step 209 may be repeated until the target centroid 210 is determined to be within a predetermined threshold distance from the old cluster centroid 207.

15 In a cluster-selection step 211, a new cluster 201 may be selected and the method may proceed with the cluster-centroid step 204. In an alternative embodiment, where the cluster-count may be adjusted, the method may proceed with the cluster-setup step 203.

20 The foregoing steps may be repeated for all clusters 201 for the selected feature field 105.

In a tree-structure step 212, a second feature field 105 may be selected and the method may proceed with the 25 cluster-setup step 203. A set of clusters 201 for the second feature field 105 may be determined for each cluster 201 for the first feature field 105, forming a second level of a tree structure 213 of clusters 201. The tree-structure step 212 may be repeated for succeeding feature fields 105 until it is performed for all 30 nonsuperfluous feature fields 105.

The use of such a tree structure 213 of clusters 201 as a tool for data analysis is well known in the art, particularly as a technique for data compression. In a preferred embodiment, a technique known as adaptive k-35 means clustering may be used to help determine the tree structure 213. Records embodying the tree structure 213 may be created as a second database 101 which incorporates

essentially the same information. Where the database 101 is predictive, cluster recognition may be used as a technique which performs data compression and maintains the new database 101 predictive.

5 Methods shown for the behavior module of copending application Serial No. _____, Lyon & Lyon Docket No. 193/304, which are applicable to cases, may also be applied to the records of a predictive database 101. In a preferred embodiment, the machine learning 10 system may use the methods shown in that copending application to determine which records are most "useful", i.e. which are good exemplar records for predicting values of f^* , by analogy to those cases which would be good exemplar cases.

15 By analogy to copending application Serial No. _____, Lyon & Lyon Docket No. 193/304, the machine learning system may tune the database 101 in several ways. It may add records 102 which are newly encountered, by analogy to adding cases which are new 20 exemplar cases. It may remove those records 102 which are least "useful", by analogy to removing cases which are poor exemplar cases. It may generate new records 119 by a genetic technique, by analogy to generating new cases by a genetic technique. It may add such new records 119 to 25 the database 101 and remove those new records 119 which fail to compete.

The machine learning system may also implement a case-based-like reasoning system with a relational database. Cases in a case-based reasoning system may be 30 represented by records 102 like those in the database 101, and the features of a case may be represented by the fields 103 of the record 102. For example, a case with two features with numerical values and one feature with a text value may be represented by a record 102 with two 35 fields 103 with numerical values and one field 103 with a text value. Cases in the case base may be represented by records 102 in the database 101, while cases which are

encountered and which may be matched to the case base may be represented by records 102 which may be matched to the database 101.

When a case is to be matched to the case base, a 5 search designation 111 may be composed and applied so as to produce a search set 114 of records 102 which represent "similar" cases. One of these records 102 may be chosen as the predictive record 116, which represents the case which is the "best match". When the record 102 which 10 represents the best match is chosen, the predicted fields 106 may represent the prescribed action for that case. For example, in a help-desk system the predicted fields 106 may indicate a voice response message and selection menu to be presented to the caller.

15 Evaluation of cases in a case-based system for accuracy and utility (or other values) may be represented by evaluating accuracy and utility (or other values) of the records 102 in the database 101.

The machine learning system may also implement an 20 autonomous learning system, like that disclosed in copending application Serial No. _____, Lyon & Lyon Docket No. 193/304, with a relational database. As shown therein, the autonomous learning system may comprise a case base in which cases are selected by a genetic 25 technique, in which cases may be generated, deleted or modified. For example, new cases may be generated by altering features of cases already in the case base.

As noted herein, cases in a case-based reasoning system may be represented by records 102 like those in the 30 database 101. The records 102 in the database 101 may also be generated, deleted or modified by means of techniques like those disclosed in copending application Serial No. _____, Lyon & Lyon Docket No. 193/304. Where those techniques generate new cases with 35 particular features, the machine learning system may generate new records 102 with fields 103 which correspond to those features and which represent those cases. Where

those techniques delete cases, the machine learning system may remove the records 102 from the database 101 which correspond to those cases. Where those techniques modify cases (by modifying particular features of those cases), 5 the machine learning system may modify records 102 which correspond to those cases (by modifying fields 103 which correspond to those features).

Appendix A (pages ____-____) shows an example software environment and autonomous agent for 10 distinguishing between classes of irises. The example software environment comprises an SQL table having a set of fields which relate to iris features, a set of SQL statements which exercise the machine learning system, and a machine learning system which makes the table 15 predictive. Some exemplary data statements are also included.

Alternative Embodiments

While preferred embodiments are disclosed herein, many variations are possible which remain within the 20 concept and scope of the invention, and these variations would become clear to one of ordinary skill in the art after perusal of the specification, drawings and claims herein.

Claims

1. A machine learning system comprising means for autonomous learning using a relational database.
2. A machine learning system comprising means for 5 performing case-based reasoning using a relational database.
3. A machine learning system, comprising means for examining a selected entry in a database; 10 means for querying said database for a set of entries which are representative of said selected entry; and means for predicting a value for at least one field of said selected entry in response to said set of 15 representative entries.
4. A system as in claim 3, wherein said database is a relational database.
5. A system as in claim 3, comprising means for repeatedly triggering said means for examining, means for 20 querying and means for predicting.
6. A system as in claim 3, comprising means for evaluating at least one of said set of representative entries.
7. A system as in claim 3, comprising means for 25 recording, for at least one of said set of representative entries, an evaluation of that entry for predicting one or more fields of said selected entry.
8. A system as in claim 7, wherein said evaluation comprises a measure of accuracy or utility.

9. Apparatus comprising

means for choosing a predictive record from among a database of records in response to an evaluation field found in at least one record in said database;

5 means for comparing a predicted value found in said predictive record with a selected actual value; and means for updating said evaluation field in said predictive record in response to said comparison.

10. Apparatus as in claim 9, wherein said means for choosing comprises

means for applying a search designation to said database to produce a search set of records; and

means for choosing a predictive record in response to an evaluation field found in at least one 15 record in said search set.

11. Apparatus as in claim 10, comprising means for of composing said search designation in response to a selected record, wherein said selected actual value is chosen in response to said selected record.

20 12. Apparatus as in claim 11, wherein said selected record is a record in said database.

13. Apparatus comprising

means for designating a selected record from among a plurality of records in a database;

25 means for composing a search designation in response to said selected record and in response to a set of similarity tables;

means for applying said search designation to said database to produce a search set of records;

30 means for choosing a predictive record in response to an evaluation field found in each record in said search set;

means for comparing a predicted value from said predictive record with an actual value from said selected record; and

means for updating said evaluation field in said predictive record.

14. Apparatus comprising

means for updating a database of records, said means for updating comprising (1) choosing a predictive record from among a database of records in response to an evaluation field found in at least one record in said database, (2) comparing a predicted value found in said predictive record with a selected actual value, and (3) updating said evaluation field in said predictive record in response to said comparison; and

15 means for repeatedly activating said means for updating until said database remains substantially unchanged.

15. Apparatus comprising

means for predicting a predicted value for at least one field of a relational database; means for comparing said predicted value with an actual value for said at least one field; and means for generating a signal in response to said comparison.

25 16. Apparatus as in claim 15, wherein said means for generating comprises means for indicating an error when said predicted value differs substantially from said actual value.

30 17. Apparatus as in claim 15, wherein said means for generating comprises means for replacing said actual value with said predicted value.

18. Apparatus as in claim 15, comprising means for indicating, for at least one actual value, that said actual value is missing data or untrustworthy data; wherein said means for comparing is responsive to said 5 means for indicating.

19. Apparatus as in claim 18, wherein said means for generating comprises means for replacing said actual value with said predicted value when said actual value is missing data or untrustworthy data.

10 20. Apparatus as in claim 18, wherein said means for generating comprises means for indicating, for at least one actual value, that said actual value is trustworthy data when said predicted value does not differ substantially from said actual value.

15 21. A method of machine learning, comprising the steps of

examining a selected entry in a database;

querying said database for a set of entries which are representative of said selected entry; and

20 predicting a value for at least one field of said selected entry in response to said set of representative entries.

22. A method as in claim 21, wherein said database is a relational database.

25 23. A method as in claim 21, comprising the step of repeatedly performing said steps of examining, querying and predicting.

24. A method as in claim 21, comprising the step of evaluating at least one of said set of representative 30 entries.

25. A method as in claim 21, comprising the step of recording, for at least one of said set of representative entries, an evaluation of that entry for predicting one or more fields of said selected entry.

5 26. A method as in claim 25, wherein said evaluation comprises a measure of accuracy or utility.

27. A method comprising the steps of
choosing a predictive record from among a
database of records in response to an evaluation field
10 found in at least one record in said database;
comparing a predicted value found in said
predictive record with a selected actual value; and
updating said evaluation field in said
predictive record in response to said comparison.

15 28. A method as in claim 27, wherein said step of
choosing comprises the steps of
applying a search designation to said database
to produce a search set of records; and
choosing a predictive record in response to an
20 evaluation field found in at least one record in said
search set.

29. A method as in claim 28, wherein said method is
repeated with a plurality of new search designations.

30. A method as in claim 28, comprising the step of
25 composing said search designation in response to a
selected record, wherein said selected actual value is
chosen in response to said selected record.

31. A method as in claim 30, wherein said selected
record is a record in said database.

32. A method as in claim 30, wherein said method is repeated with substantially all records in said database, each one in turn, being designated as said selected record.

5 33. A method comprising the steps of designating a selected record from among a plurality of records in a database; composing a search designation in response to said selected record and in response to a set of 10 similarity tables; applying said search designation to said database to produce a search set of records; choosing a predictive record in response to an evaluation field found in each record in said search set; 15 comparing a predicted value from said predictive record with an actual value from said selected record; and updating said evaluation field in said predictive record.

20 34. A method comprising the steps of repeatedly updating a database of records until said database remains substantially unchanged, wherein said step of updating comprises the steps of (1) choosing a predictive record from among a database of records in response to an evaluation field found in at least one 25 record in said database, (2) comparing a predicted value found in said predictive record with a selected actual value, and (3) updating said evaluation field in said predictive record in response to said comparison.

30 35. A method, comprising the steps of predicting a predicted value for at least one field of a relational database; comparing said predicted value with an actual value for said at least one field; and

generating a signal in response to said comparison.

36. A method as in claim 35, wherein said step of generating comprises the step of indicating an error when 5 said predicted value differs substantially from said actual value.

37. A method as in claim 35, wherein said step of generating comprises the step of replacing said actual value with said predicted value.

10 38. A method as in claim 35, comprising the step of indicating, for at least one actual value, that said actual value is missing data or untrustworthy data; wherein said step of comparing is responsive to said step of indicating.

15 39. A method as in claim 38, wherein said step of generating comprises the step of replacing said actual value with said predicted value when said actual value is missing data or untrustworthy data.

20 40. A method as in claim 38, wherein said step of generating comprises the step of indicating, for at least one actual value, that said actual value is trustworthy data when said predicted value does not differ substantially from said actual value.

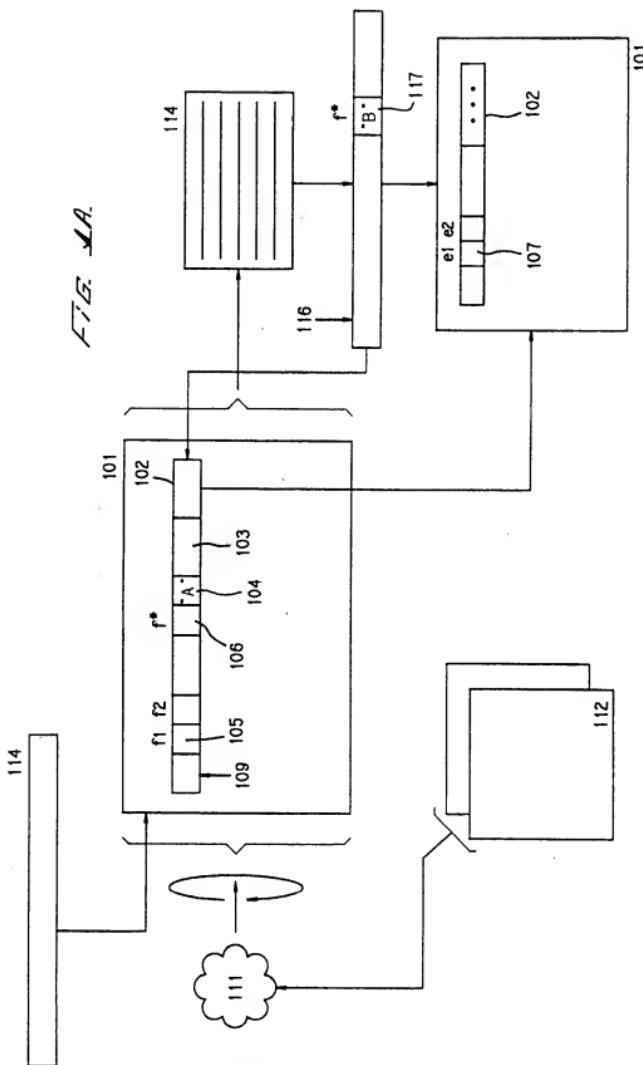
41. A predictive database.

25 42. A predictive database as in claim 41, said predictive database having been constructed by repeatedly updating a database of records until said database remains substantially unchanged.

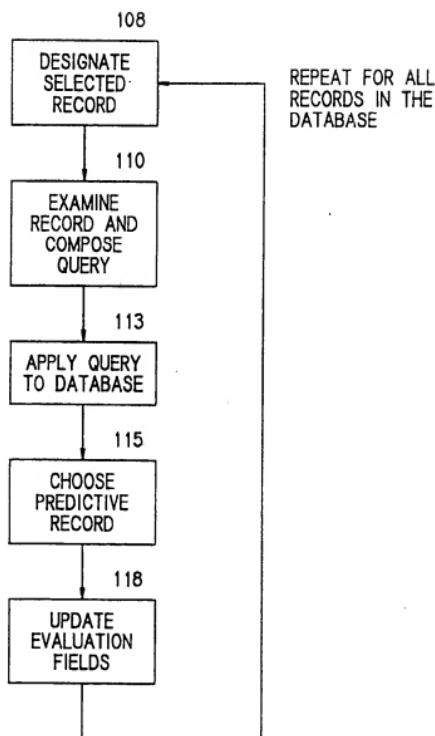
43. A predictive database as in claim 41, said predictive database having been constructed by performing, at least once, the steps of (1) choosing a predictive record from among a database of records in response to an evaluation field found in at least one record in said database, (2) comparing a predicted value found in said predictive record with a selected actual value, and (3) updating said evaluation field in said predictive record in response to said comparison.

10 44. A self-repairing database, comprising means for indicating an error in a relational database; means for replacing data indicated to be in error in a relational database; and 15 means for repeatedly triggering said means for indicating and said means for replacing.

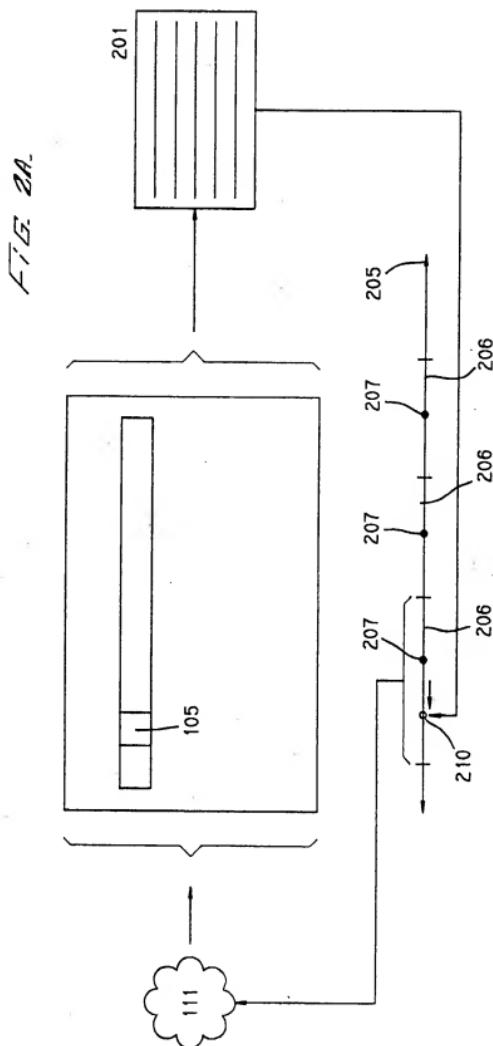
1 / 4



SUBSTITUTE SHEET



3 / 4



SUBSTITUTE SHEET

4/4

